Hotspot Management Strategy for Real-Time Log Data in MongoDB

Boonyalit Oonhawat, Natawut Nupairoj

Department of Computer Engineering, Chulalongkorn University, Bangkok, Thailand, 10330

boonyalit.o@student.chula.ac.th, natawut.n@chula.ac.th

***Abstract*—MongoDB is a Document based NoSQL that developed to answer the increasing demands for scalable date store in big data era. To achieve good performance, the system must be designed properly from the start. The performance issues like hotspot can be fatal to overall performance and various researches were focusing on applying load balancing techniques to resolve this problem. However, these techniques may not be effective for time series data, such as real-time system logs. In this paper, we propose a new data distribution algorithm based on tag aware sharding to minimize the effect of hotspot problem, especially the system with heavy writing requirements. This improves the overall performance and allow us to handle time series data effectively.**

***Keyword*—Big Data, Database, Data Distribution, Load Balance, Log Data, MongoDB, NoSQL, Sharding, Hot Spot**

Boonyalit Oonhawat graduated from Assumption University, Thailand in 2011, Bachelor degree in Business Administration and Major in Business Information System. Currently study at Chulalongkorn University Computer Engineering department. He also experienced in several computer field such as web development but currently focus on researching cloud computing, distributed system and database.

Natawut Nupairoj received the B.E. in Computer Engineering from Chulalongkorn University in 1990, M.S. and Ph.D. in Computer Science from Michigan State University in 1993 and 1998, respectively. Since 1998, he has been a Lecturer at Department of Computer Engineering, Chulalongkorn University. His research interests include grid computing, distributed system, and service oriented architecture.